

28p.  
N64-25689  
CODE-1  
nasa Cr-56789  
Cat. 08

Technical Report No. 32-602

**Properties of Error-Correcting Codes  
at Low Signal-to-Noise Ratios**

Edward C. Posner

OTS PRICE  
XEROX \$ 2.60 ph.  
MICROFILM \$ \_\_\_\_\_

001


JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

June 15, 1964

*Technical Report No. 32-602*

***Properties of Error-Correcting Codes  
at Low Signal-to-Noise Ratios***

*Edward C. Posner*

  
M. Easterling, Chief  
Communications Systems Research Section

**JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA**

June 15, 1964

Copyright © 1964  
Jet Propulsion Laboratory  
California Institute of Technology

Prepared Under Contract No. NAS 7-100  
National Aeronautics & Space Administration

## CONTENTS

I. Introduction .....	1
II. Minimizing Expected-Equivalent Bit Error Probability .....	2
III. Minimizing Word-Equivalent Bit Error Probability .....	6
IV. Orthogonal Codes.....	13
References .....	23

## FIGURES

1. Power gain vs. signal-to-noise ratio for (7, 4) code .....	4
2. Power gain vs. signal-to-noise ratio for (127, 120) code .....	5
3. $A_\nu$ vs. $\nu$ for $1 \leq \nu \leq 7$ .....	19

## ABSTRACT

25689

The use of error-correcting codes on a white Gaussian channel is considered as the signal-to-noise ratio approaches zero. Two criteria of performance are used: the expected number of information bits in error and the probability of word error. It is shown that if bit-by-bit detection is used, and if the expected number of bits in error is to be minimized, then maximum-likelihood decoding should be abandoned at low signal-to-noise ratios. In fact, coding should be abandoned altogether at low signal-to-noise ratios in favor of longer integration time per bit. If word error probability is to be minimized, then error-correcting codes using bit-by-bit detection hardly yield any gain, and usually result in a loss. However, orthogonal codes using correlation detection give a gain approaching 3.4 db as the length of the code increases without bound.

*Author*

## I. INTRODUCTION

In a recent article (Ref. 1), Hackett derived asymptotic error probabilities for error-correcting codes used on a white Gaussian channel at high signal-to-noise ratio. Another important problem is the behavior of error-correcting codes at low signal-to-noise ratios, since coding is even more necessary in such poor environments. This Report derives formulas for error probabilities of error-correcting codes at signal-to-noise ratios approaching zero. The same formulas are derived for orthogonal codes using correlation detection. The qualitative conclusion is that at low signal-to-noise ratios, the performance of orthogonal codes far surpasses the performance of error-correcting codes using bit-by-bit detection.

Before any analysis can be performed, it is necessary to review two definitions of error probabilities for binary codes used with bit-by-bit detection, since at low signal-to-noise ratios, the two probabilities can differ significantly. The discussion will be concerned mainly with bit error probabilities rather than word error probabilities, since systems with differing word length will be compared. Two kinds of bit error probability are defined:

1.  $p_W$ , "word-equivalent bit error probability,"
2.  $p_B$ , "expected-equivalent bit error probability."

The definition of  $p_W$  is as follows: if a code with  $k$  information bits is considered,  $p_W$  is that bit error probability which would result in the same word error probability, for a  $k$ -bit uncoded word, as the word error probability with the code actually used. And  $p_B$  is defined as that bit error probability which results in the same expected number of information bits in error, for  $k$  uncoded bits, as the expected number in error when the code is used.

In other words, if  $p_1$  is the probability of word error when using a given code,  $1 - (1 - p_W)^k = p_1$ . And if  $s$  is the expected number of information bits in error when the code is used, then  $k p_B = s$ .

## II. MINIMIZING EXPECTED-EQUIVALENT BIT ERROR PROBABILITY

Maximum-likelihood decoding, by definition, minimizes  $p_W$ , not  $p_B$ . If the words have some significance as *words*, then minimization of  $p_W$  is indeed the correct criterion. But if the  $k$  information bits in the code word have no particular relationship to each other, but arise from unrelated experiments, one would consider  $p_B$  the correct performance criterion for the system. It is, as will be shown below, not true that maximum-likelihood decoding always minimizes  $p_B$ .

The decoding scheme that does minimize  $p_B$  is the following procedure. For a given received word, a given information bit is decoded as a 0 if the conditional probability that the given symbol is a 0, given the received word, is greater than the conditional probability that it is a 1. This is done independently for each information bit; this procedure clearly minimizes the expected number of information bits in error.

However, this decoding scheme rarely coincides with maximum-likelihood decoding at low signal-to-noise ratios. (One case in which it actually coincides is the code in which a single information bit is repeated  $n$  times.) On the other hand, for high signal-to-noise ratios, this bit-by-bit scheme always agrees with maximum-likelihood decoding. To prove this, observe that the most likely transmitted word, given a received word, becomes arbitrarily more likely than the next-most-likely transmitted word, at sufficiently high signal-to-noise ratios. At low signal-to-noise ratios, the optimum decoding scheme to minimize  $p_B$  for many codes becomes the following: the information bits are accepted as correct, thus ignoring the check bits entirely. The asymptotic bit power gain is defined as the limit, as the signal approaches zero, of the number by which the signal must be multiplied when using the code to achieve the same  $p_B$  as one has when not using the code. This power gain is then  $k/n$  (*less than 1*) for codes in which the decoding scheme is to ignore the check bits. For, in the same word time, it would be possible to transmit the  $k$  information bits uncoded and use  $(n/k)t$  time per symbol, where  $t$  was the time used in the coded case. Since signal-to-noise ratio is proportional to symbol integration time,  $n/k$  times as much signal power would be obtained by using "bit expansion."

It is conjectured that with any code (linear or otherwise), except for the trivial "no-coding" code, the asymptotic power gain using bit-by-bit detection, even with the optimum bit-by-bit decoding scheme for minimizing  $p_B$ , is less than 1.

Now the asymptotic bit power gain obtained with maximum-likelihood decoding will be considered. We can prove that for close-packed cyclic codes, such as the Golay (23, 12) code, the asymptotic bit power gain using maximum-likelihood decoding is  $(k/n) [1 - (2\gamma/n)]^2$ , where  $\gamma$  is the average weight of the correctable errors. Thus maximum-likelihood decoding is worse by a factor of  $[1 - (2\gamma/n)]^2$  than the scheme that ignores the check bits. The proof is omitted. The power gain vs. signal-to-noise ratio for the Hamming (7, 4) and (127, 120) codes, using maximum-likelihood decoding, is graphed in Figs. 1 and 2. It may be noted that the crossover point, below which the codes should be abandoned, occurs at signal-to-noise ratios that are to be considered rather high.

For criterion  $p_B$ , then, the question is essentially answered. Ultimately a given code with any decoding scheme whatsoever should be abandoned as the signal-to-noise ratio approaches zero. The time saved should then be used to expand the information bit time by  $n/k$ . The effect of this result is that as the signal-to-noise ratio approaches zero, longer and longer codes must be used to achieve any gain at all with criterion  $p_B$ . However, the use of orthogonal (or even better, biorthogonal) codes with correlation detection is known to always decrease  $p_B$  (Viterbi, Ref. 2). The quoted result, coupled with the discussion above, shows that orthogonal codes with correlation detection are the best choice at low signal-to-noise ratios, at least for criterion  $p_B$ .



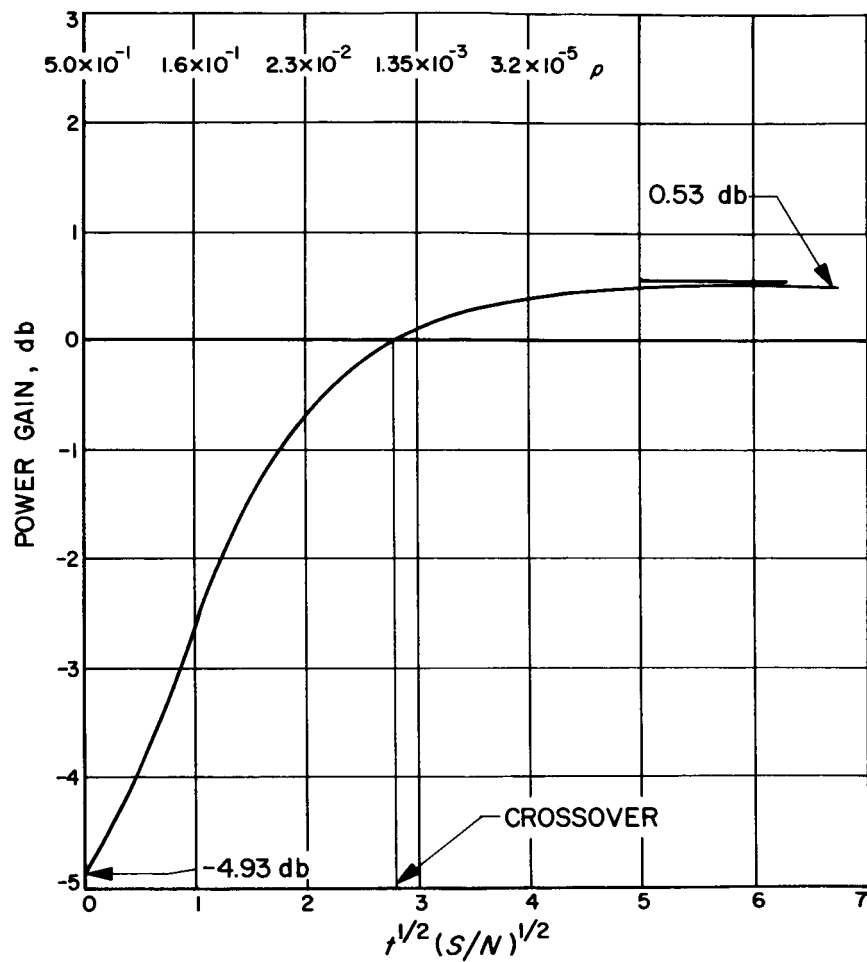


Fig. 1. Power gain vs. signal-to-noise ratio for (7, 4) code

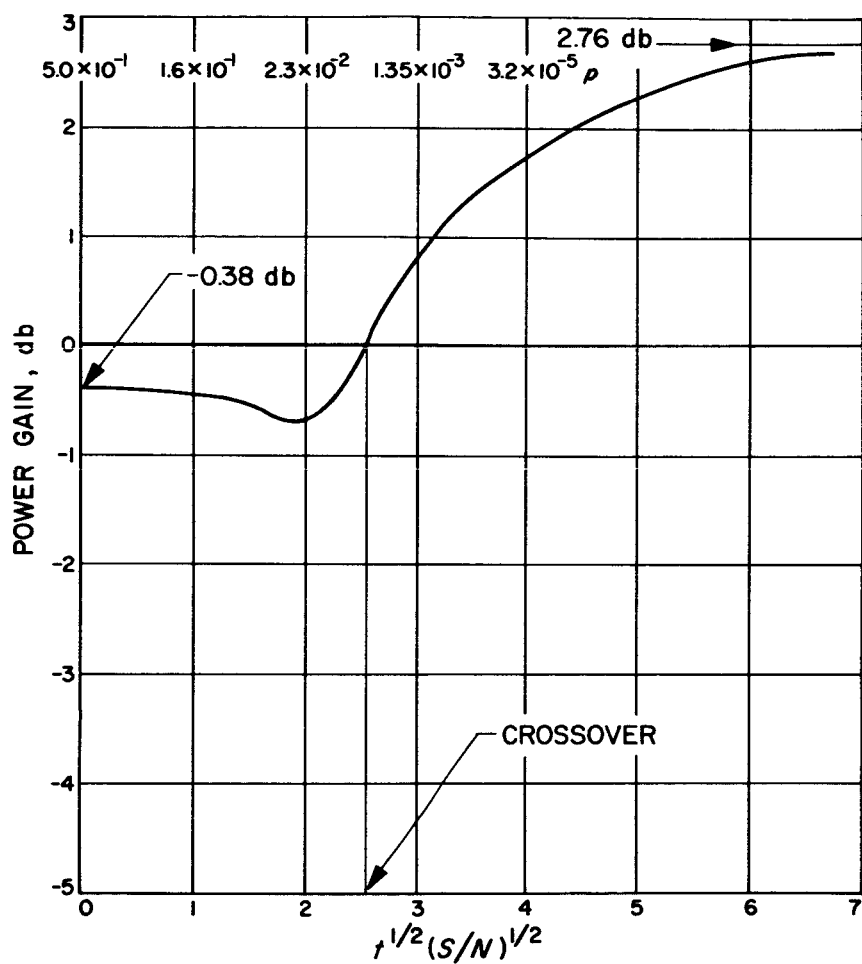


Fig. 2. Power gain vs. signal-to-noise ratio for (127, 120) code

### III. MINIMIZING WORD-EQUIVALENT BIT ERROR PROBABILITY

The question of the asymptotic utility of a fixed error-correcting code at signal-to-noise ratios approaching zero is completely answered in the fashion described above, when the criterion of minimizing  $p_B$  is used. Henceforth, we treat the criterion  $p_W$  (used, of course, with maximum-likelihood decoding). We can then revert to word error probabilities to compare coding schemes with the same  $k$ . The situation is not as clear-cut with the  $p_W$  criterion as it was with the  $p_B$  criterion, and some error-correcting codes actually do improve performance even at arbitrarily low signal-to-noise ratios. Nevertheless, orthogonal codes with correlation detection do even better, as will be shown.

An expression for the probability of correct word reception will be obtained for a given error-correcting code used with maximum-likelihood detection, asymptotically valid as the signal-to-noise ratio approaches zero. (The case  $n = k$  is the no-coding case.) The discussion will be restricted for simplicity to linear cyclic codes. The parameter that will approach zero is the integration time per bit,  $t$ ; the signal-to-noise ratio per bit is proportional to  $t$ .

Define  $p(t)$  to be the individual uncoded bit error probability for integration time  $t$ . The symbols are considered to be  $\pm 1$  for this analysis. The variance of the integrator output for  $t = 1$  is normalized to be 1. Then an individual bit is received as  $+1$  if the integral of the signal over 1 second is positive; similarly,  $-1$  if that integral is negative. The random variable of the additive noise component of this integral is Gaussian of mean 0 and variance  $t$ . The probability of symbol error,  $p(t)$ , is thus the probability that a Gaussian variable of mean 0 and variance  $t$  exceeds 1. Thus

$$p(t) = 1 - \int_{v=-\infty}^{t^{1/2}} (2\pi)^{-1/2} \exp(-v^2/2) dv = 1 - \Phi(t^{1/2}), \quad (1)$$

where  $\Phi$  is the cumulative unit normal distribution function.

Various coding systems will be compared with the no-coding system. The change in power necessary to produce the same word error probability will now be derived. The use of an  $(n, k)$  code such that the code word occupies the same time ( $kt$ ) with the code as without it decreases the integration time per symbol in the coded case by a factor of  $k/n$ , since the time available per  $k$ -bit word is the same whether a code is used

or not. Thus the symbol error probability using coding,  $r(t)$  say, becomes

$$\begin{aligned} r(t) &= 1 - \int_{v=-\infty}^{(kt/n)^{1/2}} (2\pi)^{-1/2} \exp(-v^2/2) dv \\ &= 1 - \Phi\left[\left(\frac{kt}{n}\right)^{1/2}\right]. \end{aligned} \quad (2)$$

The following notation will be used:

$$p(t) = \frac{1}{2} - \delta,$$

$$\delta = \int_{v=0}^{t^{1/2}} (2\pi)^{-1/2} \exp(-v^2/2) dv = \frac{1}{2} - \Phi(t^{1/2});$$

$$q(t) = 1 - p(t),$$

$$r(t) = \frac{1}{2} - \epsilon,$$

$$\epsilon = \int_{v=0}^{(kt/n)^{1/2}} (2\pi)^{-1/2} \exp(-v^2/2) dv = \frac{1}{2} - \Phi\left[\left(\frac{kt}{n}\right)^{1/2}\right];$$

$$k(t) = 1 - r(t).$$

Now suppose there is given a linear cyclic code  $C$  of length  $n$  with  $k$  information bits, and let maximum-likelihood decoding be used. Let  $f_j$ ,  $0 \leq j \leq n$ , denote the number of error patterns of weight  $j$  corrected by the code  $C$ . It is well known that  $\sum_{j=0}^n f_j = 2^{n-k}$ ; for proof, observe that  $2^{n-k}$  is the number of cosets of  $C$  in the space of binary  $n$ -tuples. This result is used presently.

The probability that a  $k$ -bit uncoded word is received correctly, say  $1 - p_{WU}$ , is given by

$$1 - p_{WU} = [q(t)]^k. \quad (3)$$

The probability that the  $n$ -bit coded word is decoded correctly using  $C$ , say  $1 - p_{WC}$ , is given by

$$1 - p_{WC}(t) = \sum_{j=0}^n f_j [r(t)]^j [s(t)]^{n-j}. \quad (4)$$

To prove Eq. (4), observe that the transmitted word is the output of the maximum-likelihood receiver if and only if a correctable error pattern has been added to the transmitted word. Equation (4) merely represents the probability of making a correctable error.

In the following theorem, we shall need the quantity  $\gamma = \gamma(C)$ , the "average weight of a correctable error":

$$\gamma = \left(\frac{1}{2}\right)^{n-k} \sum_{j=0}^n j f_j.$$

The parameter  $\gamma$  is difficult to compute for most codes, since it depends on the weight structure. In the example to be given after Theorem 1,  $\gamma$  will, however, be computed for the class of close-packed single-error-correcting binary Hamming Codes.

Next define  $G(t)$ , the power gain ratio obtained by using  $C$ , to be that factor by which the integration time per word, or, equivalently, energy per bit, in the no-coding case would have to be multiplied in order to have the same probability of correct reception with the use of the code as without the code, when the word time is  $kt$ . (If  $G(t) < 1$ , there is a power loss if  $C$  is used.) Define  $G [= G(C)]$  as  $G = \lim_{t \rightarrow 0} G(t)$ , the limiting power gain as the signal-to-noise ratio approaches zero. The following theorem can then be stated:

$$\text{Theorem 1: } G = \frac{(n - 2\gamma)^2}{nk}.$$

Before embarking on the proof of Theorem 1, the following corollary will be noted.

*Corollary:* The limiting power gain using  $C$  exceeds 1 if and only if  $\gamma/n < (1/2)(1 - \sqrt{k/n})$ , that is, if and only if the average density of correctable errors is less than half of 1 minus the square root of the information rate  $k/n$ .

This corollary is reasonable, since for given  $n, k$ , the better codes are those that correct the greatest number of more likely error patterns, i.e., error patterns of small weight.

To prove Theorem 1, we first show that

$$1 - p_{WC}(t) = \frac{1}{2^k} [1 + 2\epsilon(n - 2\gamma) + O(\epsilon^2)], \text{ as } t \rightarrow 0. \quad (5)$$

To prove Eq. (5), Eq. (4) is written as

$$\begin{aligned} \sum_{j=0}^n f_j \left[ \left( \frac{1}{2} \right) - \epsilon \right]^j \left[ \left( \frac{1}{2} \right) + \epsilon \right]^{n-j} &= \\ \sum_{j=0}^n f_j \left[ \frac{1}{2^j} - \frac{j\epsilon}{2^{j-1}} \right] \left[ \frac{1}{2^{n-j}} + \frac{(n-j)\epsilon}{2^{n-j-1}} \right] + O(\epsilon^2) &= \\ \sum_{j=0}^n f_j \left[ \frac{1}{2^n} - \frac{j\epsilon}{2^{n-1}} + \frac{(n-j)\epsilon}{2^{n-1}} \right] + O(\epsilon^2) &= \\ \frac{1}{2^n} \sum_{j=0}^n f_j + \frac{n\epsilon}{2^{n-1}} \sum_{j=0}^n f_j - \frac{2\epsilon}{2^{n-1}} \sum_{j=0}^n jf_j + O(\epsilon^2). \end{aligned}$$

Using the fact that  $\sum_{j=0}^n f_j = 2^{n-k}$  and also the definition of  $\gamma$ , Eq. (5) is proved.

For the no-coding case, we have more simply from Eq. (3)

$$1 - p_{WU}(t) = \frac{1}{2^k} (1 - 2k\delta) + O(\delta^2). \quad (6)$$

To find  $\lim_{t \rightarrow 0} G(t)$ , a relationship between  $\epsilon$  and  $\delta$  must be found. As  $t$  approaches zero, we have from the definitions of  $\epsilon$  and  $\delta$ , using the power series for  $e^{-v^2/2}$  and integrating term-by-term, that

$$\begin{aligned}\delta &= \left( \frac{t^{1/2}}{\sqrt{2\pi}} \right) [1 + O(t)] , \\ \epsilon &= \left[ \frac{\left( \frac{kt}{n} \right)^{1/2}}{\sqrt{2\pi}} \right] [1 + O(t)] .\end{aligned}\tag{7}$$

Thus Eqs. (5) and (6) become

$$1 - p_{WC}(t) = \frac{1}{2^k} \left[ 1 + \frac{2}{\sqrt{2\pi}} \left( \frac{kt}{n} \right)^{1/2} (n - 2\gamma) + O(t) \right]$$

and

$$1 - p_{WU}(t') = \frac{1}{2^k} \left[ 1 - \frac{2k(t')^{1/2}}{\sqrt{2\pi}} + O(t') \right] .\tag{8}$$

The notation  $t'$  is used in Eq. (8) to indicate that the integration time  $kt$  per  $k$  symbol word when using the code is to be different from the time  $kt'$  per  $k$  symbol word when not using the code.

We now set  $1 - p_{WC}(t) = 1 - p_{WU}(t')$  and see what relationship between  $t$  and  $t'$  this implies.

Equating Eqs. (7) and (8), one has

$$\left( \frac{kt}{n} \right)^{1/2} (n - 2\gamma) + O(t) = -k(t')^{1/2} + O(t') ,\tag{9}$$

or

$$(t')^{1/2} = t^{1/2} (nk)^{-1/2} (2\gamma - n) + O[\max(t, t')] ,\tag{10}$$

or

$$t' = \frac{(n - 2\gamma)^2}{nk} t + O\{\max[t^{3/2}, (t')^{3/2}]\} .\tag{11}$$

Equation (11) means that  $G(t) = (n - 2\gamma)^2/nk + O\{\max[t^{1/2}, (t')^{1/2}]\}$ . Since  $t'$  approaches zero as  $t$  approaches zero,  $\lim_{t \rightarrow 0} G(t) = (n - 2\gamma)^2/nk$ , which completes the proof of Theorem 1.

The example of the close-packed single-error-correcting Hamming codes (Ref. 3, Sec. 5.1) will now be treated. For any  $m \geq 2$ , these Hamming codes have  $n = 2^m - 1$ ,  $k = 2^m - m - 1$ , and have minimum distance 3. The correctable error patterns are precisely the zero error pattern and every error pattern of weight 1. Thus  $\gamma = (1/2^{n-k}) [(2^{n-k} - 1) \cdot 1 + 1 \cdot 0]$ , and  $\gamma = 1 - 2^{-m}$ . For these Hamming codes, indexing  $G$  by the parameter  $m$ , Theorem 1 now gives  $G_m = (2^m - 3 + 2^{-m+1})^2 / (2^m - 1)(2^m - m - 1)$ , which converges to 1 as  $m \rightarrow \infty$ . For the (7, 4) code,  $G_3 = 63/64$  is less than 1. However,  $G_m$  is greater than 1 for  $m \geq 5$ ;  $G_m$  increases up to  $m = 5$ , then decreases toward 1.

Now let us consider the Golay (23, 12) triple-error-correcting code (Ref. 3, p. 70). Here  $n = 23$  and  $k = 12$ ; since the code is close packed, that is, since the code corrects all errors of weight  $\leq 3$  and no others, we can compute  $\gamma$  as follows:

$$\gamma = \left[ 1 \cdot 0 + 23 \cdot 1 + \binom{23}{2} \cdot 2 + \binom{23}{3} \cdot 3 \right] 2^{-11} = 2.858 \dots$$

Thus  $G = (23 - 5.716)^2 (23 \cdot 12)^{-1} = 1.037$ , which exceeds 1. Hence there is a gain with the Golay code, however extremely slight, even at arbitrarily low signal-to-noise ratios, if  $p_W$  is the criterion.

In fact, we can conclude from results in the next Section that  $G$  is uniformly bounded for all  $C$ . Moreover, we shall see that even with the use of transorthogonal codes using the optimal detection scheme rather than bit-by-bit detection, the asymptotic savings ratio is bounded by  $\pi \log 2 \approx 2.18$ . *A fortiori*, then,  $\pi \log 2$  is a bound for *all* error-correcting codes using the suboptimal bit-by-bit detection. The exact bound would be much less. For example, if  $k = 1$ ,  $G$  approaches  $2/\pi$  as  $n$  approaches infinity. We are unable to obtain a uniform upper bound for error-correcting codes better than  $\pi \log 2$ . We can however, prove the following: if the number of errors corrected,  $e$ , is less than or equal to  $n/32$ , then  $G < (15/16)^2 / [1 - H(1/16)]$ , where  $H$  is the entropy function to the base 2; thus,  $G < 4/3$ . The details are omitted. Similarly, if  $e \leq n/A$ , then  $G < 1 + o(A)$  for large  $A$ . These results are, however, much too weak. We conjecture, in fact, that there are only finitely many codes with  $G > 1$ , besides the close-packed Hamming single-error-correcting codes. (In fact, we only know of the Golay code.) We also conjecture that the value of  $G$  for the Hamming code with  $m = 5$ ,  $G_5 \approx 1.054$ , is an absolute upper bound to  $G$  for error-correcting codes using bit-by-bit detection. We remark that the truth of the conjecture that there are only finitely many codes with  $G \geq 1$  implies the



conjecture that there are only finitely many close-packed codes other than the Hamming codes and the codes with  $k = 1$ . The proof of this implication is omitted. This completes the analysis of the use of bit-by-bit detection and error-correcting codes at arbitrarily low signal-to-noise ratios.

## IV. ORTHOGONAL CODES

We now discuss orthogonal codes  $D$  using word correlation detection at very low signal-to-noise ratios. Again both criteria  $p_B$  and  $p_W$  should be considered, but, as in the bit-by-bit case, we can still conclude *a priori* that under the  $p_B$  criterion, there is again a power loss for a fixed orthogonal code as the signal-to-noise ratio approaches zero. Relying on formulas developed by Viterbi in Ref. 2, we consider codes  $D$  with  $n$  information bits, thus with  $2^n$  words, of length  $2^n$ . Instead of integration time per symbol,  $t$ , it is more convenient here to use the parameter  $s$ , the (output) signal-to-noise power ratio per bit. The relationship is  $s^2 = t$ .

In this notation, Eq. (9) of Ref. 2 becomes

$$p_{WD}(s) = 1 - \int_{v=-\infty}^{\infty} (2\pi)^{-1/2} \exp(-v^2/2) [\Phi(v) + n^{1/2}s]^{2^n-1} dv. \quad (12)$$

As  $s$  approaches zero, it is known that  $p_{WD}(s)$  approaches  $1 - 2^{-n}$ . This can be proved from Eq. (12) by integrating by parts, but a more elegant proof is given by observing that the capacity of this channel approaches zero as  $s$  approaches zero. Since  $D$  has  $2^n$  words,  $1 - p_{WD}(s)$  indeed approaches  $2^{-n}$ , as required.

We must now find the coefficient of  $s$  in the power series expansion of  $p_{WD}(s)$  around  $s = 0$ .

Equation (12) can be written

$$p_{WD}(s) = 1 - \int_{v=-\infty}^{\infty} (2\pi)^{-1/2} \exp(-v^2/2) [\Phi(v) + n^{1/2}s\phi(v)]^{2^n-1} dv + O(s^2), \quad (13)$$

where  $\phi(v) = d\Phi(v)/(dv) = (2\pi)^{-1/2} \exp(-v^2/2)$  is the unit normal density function. Using the binomial expansion,

$$p_{WD}(s) = 1 - \int_{v=-\infty}^{\infty} \{ \phi(v) [\Phi(v)^{2^n-1}] + (2^n - 1) [n^{1/2}s \phi(v)] [\Phi(v)]^{2^n-2} \} dv + O(s^2). \quad (14)$$

The value of  $1 - \int_{v=-\infty}^{\infty} \phi(v) [\Phi(v)]^{2^n-1} dv$  is  $1 - 2^{-n}$ , as derived previously. Letting  $2^n - 1 = \nu$ , we now have

$$p_{WD}(s) = 1 - 2^{-n} - n^{1/2} \nu (2\pi)^{-1/2} s \int_{v=-\infty}^{\infty} \phi(\sqrt{2v}) [\Phi(v)]^{\nu-1} dv + O(s^2). \quad (15)$$

We therefore need the value of

$$A_\nu = \int_{v=-\infty}^{\infty} \phi(\sqrt{2v}) [\Phi(v)]^{\nu-1} dv, \quad (16)$$

for  $\nu = 1, 2, \dots$ . We shall determine  $A_\nu$  exactly for  $\nu = 1, 2, 3, 4$ ; in addition, an asymptotic expression for  $A_\nu$ , valid as  $\nu$  approaches infinity, will be derived.

To this end, define for  $\alpha$  real,

$$g_\nu(\alpha) = \int_{v=-\infty}^{\infty} \phi(\sqrt{2v}) [\Phi(\alpha v)]^{\nu-1} dv; \quad (17)$$

$g_\nu(\alpha)$  is a continuously differentiable function of  $\alpha$ . We require  $g_\nu(1) = A_\nu$ ; we shall find  $g_\nu(\alpha)$  by obtaining a differential equation and then solving it. We need  $g_\nu(\alpha)$  at only the one point,  $\alpha = 1$ , but we can get  $g_\nu(\alpha)$  at three points in advance, use any one of them as a boundary condition, and check our work by seeing that the other two conditions are satisfied.

Now  $g_\nu(0) = 2^{-(\nu-1/2)}$  is immediate;  $g_1(\alpha)$  is in fact the constant function  $2^{-1/2}$ . Let  $\nu > 1$ ; then  $g_\nu(\infty) = \lim_{\alpha \rightarrow \infty} g_\nu(\alpha)$  exists, and  $g_\nu(\alpha) = \int_{v=0}^{\infty} (2\pi)^{-1/2} \exp(-v^2) dv$ . To prove the latter formula, note that  $\Phi(\alpha v)$  converges uniformly to the characteristic function of the positive axis, outside any interval containing 0, as  $\alpha$  approaches infinity. Thus  $g_\nu(\infty) = 2^{-3/2}$ ,  $\nu > 1$ . Similarly,  $g_\nu(\sqrt{2})$  can also be obtained *a priori* as  $(\sqrt{2\nu})^{-1}$ .

A first-order differential equation for  $g_\nu(\alpha)$  will be obtained in terms of  $g_{\nu-2}[(1 + \alpha^2)^{-1/2}]$  for  $\nu > 2$ ;  $g_2(\alpha)$  is determined separately. We already know that  $g_1(\alpha) = 2^{-1/2}$ , so that all  $g_\nu(\alpha)$  are ultimately determined in this manner.

It is easy to justify

$$\frac{d}{d\alpha} g_\nu(\alpha) = \frac{\nu-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp \{-v^2 [1 + (\alpha^2/2)]\} [\Phi(\alpha v)]^{\nu-2} v dv, \quad \nu \geq 2. \quad (18)$$

When  $\nu = 2$ , the  $[\Phi(\alpha v)]^{\nu-2}$  term is 1, and  $(d/d\alpha) [g_2(\alpha)]$  is the integral of the *odd* function  $(2\pi)^{-1/2} \times \exp \{-v^2 [1 + (\alpha^2/2)]\} v$  from  $-\infty$  to  $\infty$ ; hence  $(d/d\alpha) [g_2(\alpha)]$  is 0. Therefore  $g_2(\alpha)$  is a constant function, and  $g_2(\alpha) = g_2(0) = 2^{-3/2}$ .

For  $\nu > 2$ , integrate Eq. (18) by parts using  $[\Phi(\alpha v)]^{\nu-2}$  as the  $U$  in  $\int U dV = UV - \int V dU$ . After several intermediate steps, Eq. (18) yields

$$\frac{d}{d\alpha} [g_\nu(\alpha)] = \frac{(\nu-1)(\nu-2)}{2\pi} \frac{\alpha}{2+\alpha^2} \int_{-\infty}^{\infty} \phi[\sqrt{2}(1+\alpha^2)^{1/2}v] [\Phi(\alpha v)]^{\nu-3} dv, \quad \nu \geq 3. \quad (19)$$

Now define  $w = v(1 + \alpha^2)^{1/2}$  to prove

$$\frac{d}{d\alpha} g_\nu(\alpha) = \frac{(\nu-1)(\nu-2)}{2\pi} \frac{\alpha}{(2+\alpha^2)(1+\alpha^2)^{1/2}} g_{\nu-2} \left[ \frac{\alpha}{(1+\alpha^2)^{1/2}} \right], \quad \nu \geq 3. \quad (20)$$

Now  $g_3(\alpha)$  and  $g_4(\alpha)$  will be determined; the higher  $g_\nu$  are not expressible in closed form. For  $\nu = 3$ , one uses  $g_1(\alpha) = 2^{-1/2}$  and integrates Eq. (20), using  $g_3(0) = 2^{-5/2}$ , to obtain

$$g_3(\alpha) = \frac{1}{\pi\sqrt{2}} \arctan [(1 + \alpha^2)^{1/2}]. \quad (21)$$

Putting  $\alpha = 1$  in Eq. (21), one finally obtains

$$A_3 = \frac{1}{\pi\sqrt{2}} \arctan \sqrt{2} = 0.215 \dots \quad (22)$$

Using  $g_2(\alpha) = 2^{-3/2}$ , one similarly finds

$$g_4(\alpha) = \frac{3}{2\pi\sqrt{2}} \arctan(1 + \alpha^2)^{\frac{1}{2}} - \frac{1}{4\sqrt{2}} \quad (23)$$

and

$$A_4 = \left( \frac{3}{2\pi\sqrt{2}} \arctan \sqrt{2} \right) - \frac{1}{4\sqrt{2}} = 0.146 \dots \quad (24)$$

Since  $\nu = 2^n - 1$ ,  $\nu$  increases rapidly in the formulas derived for orthogonal codes. Thus, a formula for  $A_\nu$  asymptotic in  $\nu$  as  $\nu$  approaches infinity would be useful. Such an expression will now be obtained. Equation (16) can be integrated by parts to obtain

$$A_\nu = \frac{\sqrt{2\pi}}{(\nu + 1)(\nu)} \left\{ \int_{\nu=-\infty}^{\infty} \nu d([\Phi(\nu)]^{\nu+1}) \right\}; \quad (25)$$

$$A_\nu = \frac{\sqrt{2\pi}}{(\nu + 1)(\nu)} \lim_{T \rightarrow \infty} \left\{ T - \int_{\nu=-T}^T [\Phi(\nu)]^{\nu+1} d\nu \right\}, \quad (26)$$

where the limit must be introduced before integrating by parts to get from Eq. (25) to Eq. (26). We have used  $\lim_{T \rightarrow \infty} T(1 - \Phi(T)) = 0$  and  $\lim_{T \rightarrow \infty} T\Phi(T) = 0$ . From Eq. (26), using the fact that  $\Phi(-\nu) = 1 - \Phi(\nu)$ , we obtain

$$A_\nu = \frac{\sqrt{2\pi}}{(\nu + 1)(\nu)} \lim_{T \rightarrow \infty} \left\{ \int_{\nu=0}^T \{1 - [\Phi(\nu)]^{\nu+1} + [1 - \Phi(\nu)]^{\nu+1}\} d\nu \right\}. \quad (27)$$

The limit notation can now be removed:

$$A_\nu = \frac{\sqrt{2\pi}}{(\nu + 1)(\nu)} \int_{\nu=0}^{\infty} \{1 - [\Phi(\nu)]^{\nu+1} + [1 - \Phi(\nu)]^{\nu+1}\} d\nu. \quad (28)$$

We are at last prepared to derive an asymptotic expression for  $A_\nu$ . First observe that the contribution to  $A_\nu$  from the integral of  $[1 - \Phi(\nu)]^{\nu+1}$  is arbitrarily small, as  $\nu$  approaches infinity, compared with the

contribution from  $1 - [\Phi(y)]^{\nu+1}$ . Thus one has the asymptotic expression

$$A_\nu \sim \frac{\sqrt{2\pi}}{\nu^2} \int_{y=0}^{\infty} \{1 - [\Phi(y)]^{\nu+1}\} dy. \quad (29)$$

Given  $\delta$ ,  $0 < \delta < 1$ , close to 1, fixed, let  $\nu$  be so large that  $\delta < 1 - (1/2^{\nu+1})$ . Consider the set of  $y \geq 0$  with  $1 - [\Phi(y)]^{\nu+1} > \delta$ . This set is an interval to the right of 0 with the right-hand endpoint  $y_0 = y_0(\delta; \nu)$  given by  $\Phi(y_0) = (1 - \delta)^{1/(\nu+1)}$ . As  $\nu$  approaches infinity, Ref. 4, p. 106, Lemma 2, implies that

$$1 - \Phi(y_0) - \frac{1}{y_0} \phi(y_0) = O \left[ \frac{1}{y_0^3} \phi(y_0) \right] \quad (30)$$

for  $\delta$  fixed. Therefore, an expression for  $y_0$  can be obtained by setting  $1 - \Phi(y_0) = 1 - (1 - \delta)^{1/(\nu+1)}$  in Eq. (30). Now  $(1 - \delta)^{1/(\nu+1)} = \exp \{ [\log (1 - \delta)] / (\nu + 1) \} = 1 + [\log (1 - \delta)] / (\nu + 1) + O \{ [\log (1 - \delta)] / (\nu + 1) \}^2$ , for  $\delta$  fixed, as  $\nu$  approaches infinity. Thus Eq. (30) implies

$$\frac{1}{y_0} \phi(y_0) = \frac{-\log (1 - \delta)}{\nu + 1} + O \left[ \frac{1}{y_0^3} \phi(y_0) \right]. \quad (31)$$

Taking logs of both sides in Eq. (31), one has

$$\frac{-y_0^2}{2} - \log (\sqrt{2\pi} y_0) = \log \log [(1 - \delta)^{-1}] - \log (\nu + 1) + O \left[ \frac{1}{y_0^3} \phi(y_0) \right]. \quad (32)$$

Equation (32) becomes

$$\frac{y_0^2}{2} = \log \nu + O(\log y_0), \quad (33)$$

or

$$\gamma_0 = \sqrt{2 \log \nu} + O(\log \log \nu). \quad (34)$$

Equation (29) can then be written as

$$A_\nu = \frac{\sqrt{2\pi}}{\nu^2} \int_{y=0}^{\gamma_0} \{1 - [\Phi(y)]^{\nu+1}\} dy + \frac{\sqrt{2\pi}}{\nu^2} \int_{y=\gamma_0}^{\infty} \{1 - [\Phi(y)]^{\nu+1}\} dy. \quad (35)$$

For  $\delta$  fixed, the second integral is arbitrarily small when compared with the first integral, as  $\nu$  approaches infinity, using Eq. (34). Thus

$$A_\nu \sim \frac{\sqrt{2\pi}}{\nu^2} \int_{y=0}^{\gamma_0} \{1 - [\Phi(y)]^{\nu+1}\} dy. \quad (36)$$

On  $(0, \gamma_0)$ , however,  $1 - (1/2^{\nu+1}) \geq 1 - [\Phi(y)]^{\nu+1} \geq \delta$ , so that  $\int_{y=0}^{\gamma_0} \{1 - [\Phi(y)]^{\nu+1}\} dy$  lies between  $[1 - (1/2^{\nu+1})] \gamma_0$  and  $\delta \gamma_0$ . Since  $\gamma_0$  itself is asymptotic to  $\sqrt{2 \log \nu}$  for  $\delta$  fixed,  $\int_{y=0}^{\gamma_0} \{1 - [\Phi(y)]^{\nu+1}\} dy$  is asymptotic to  $\sqrt{2 \log \nu}$ . Thus

$$A_\nu \sim \frac{2}{\nu^2} \sqrt{\pi \log \nu}, \quad (37)$$

the required expression. Figure 3 is a graph of the exact  $A_\nu$  vs.  $\nu$  for  $1 \leq \nu \leq 7$ .

Equation (15) can now be written as

$$p_{WD}(t) = 1 - \frac{1}{2^n} - \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} (2^n - 1) A_{(2^n - 1)} s + O(s^2). \quad (38)$$

In the uncoded case, there are  $n$  bits which occupy time  $t$  each. Using  $r(t')$  for the probability of symbol error in this system, the probability that all  $n$  uncoded symbols are received correctly,  $p'_{WD}(t')$  say, is

$$p'_{WD}(t') = 1 - \frac{1}{2^n} - \frac{n}{2^n - 1} \left(\frac{t'}{2\pi}\right)^{\frac{1}{2}} + O(t'). \quad (39)$$

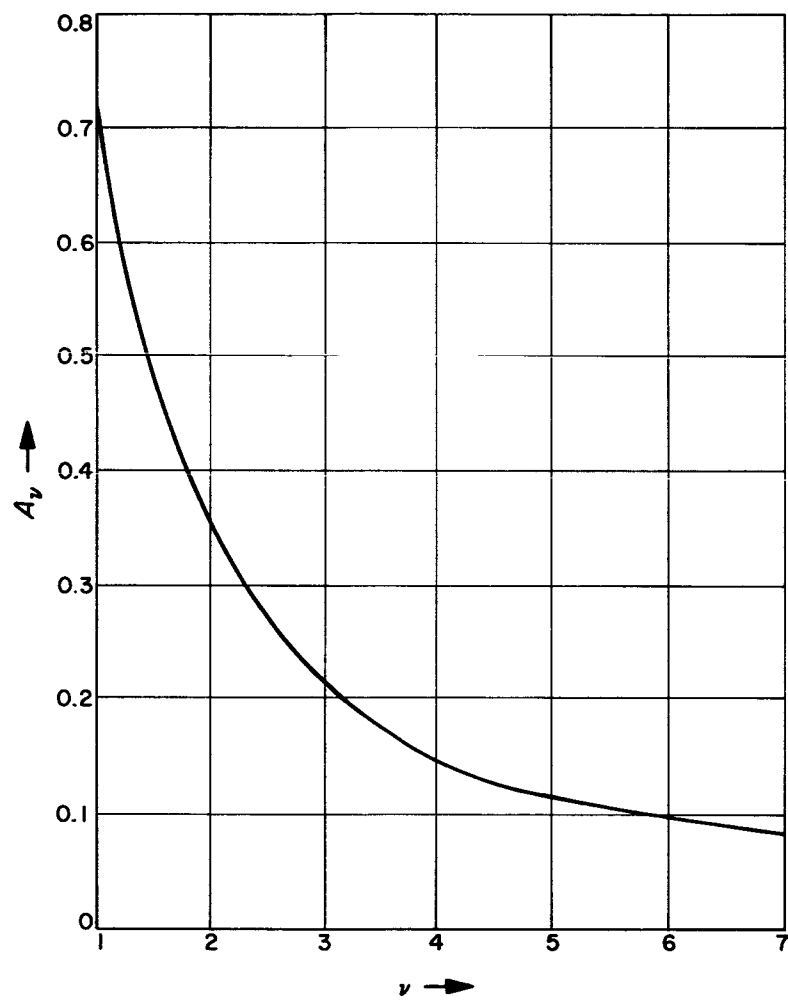


Fig. 3.  $A_\nu$  vs.  $\nu$  for  $1 \leq \nu \leq 7$



Equations (38) and (39) are equated to find a relation between  $t$  and  $t'$ :

$$\left(\frac{nt}{2\pi}\right)^{1/2} (2^n - 1) A_{2^n-1} = \frac{n}{2^n - 1} \left(\frac{t'}{2\pi}\right)^{1/2} + O[\max(t, t')], \quad (40)$$

or

$$t' = \left(\frac{t}{n}\right) 2^{2(n-1)} (2^n - 1)^2 A_{2^n-1}^2 + O\{\max[t^{3/2}, (t')^{3/2}]\}. \quad (41)$$

Thus the asymptotic power gain as  $t$  approaches zero,  $G_n(t)$  say, is

$$G_n = \frac{2^{2(n-1)} (2^n - 1)^2}{n} A_{2^n-1}^2. \quad (42)$$

For  $n = 1$ ,  $G_1 = 1/2 < 1$ , so there is a loss if an orthogonal code is used instead of no coding. For  $n = 2$ ,  $G_2 = 18 A_3^2 = (9/\pi^2) (\arctan \sqrt{2})^2 = .83 < 1$ . The reason that there is a loss for  $n = 1$  and  $n = 2$  is that in these low-dimensional cases, not coding is equivalent to using a biorthogonal code, which is better than an orthogonal code. But as  $n \rightarrow \infty$ , the difference in power gains using orthogonal as opposed to biorthogonal codes approaches zero. In addition, for  $n \geq 3$ , it can be shown that  $G_n > 1$ .

In Eq. (42), letting  $n$  approach infinity, Eq. (37) can be used to conclude that

$$\lim_{n \rightarrow \infty} G_n = \pi \log 2 \approx 2.18 \dots \quad (43)$$

Thus there is indeed a gain as  $n$  approaches infinity, the gain approaching 3.4 db. These results are summarized in the following theorem.

*Theorem 2:* For an orthogonal code with  $2^n$  words used with correlation detection on the white Gaussian channel, with word error probability criterion, the asymptotic power gain  $G_n$  is given by  $G_n = \{[2^{2(n-1)}(2^n - 1)^2] / n\} A_{2^n-1}^2$ , where  $A_\nu$ ,  $\nu \geq 1$ , is given by:  $A_1 = 1/\sqrt{2}$ ,  $A_2 = 1/(2\sqrt{2})$ ,  $A_3 = 1/(\pi\sqrt{2}) \arctan \sqrt{2} = 0.215 \dots$ ,  $A_4 = (3\sqrt{2})/4\pi (\arctan \sqrt{2}) - 1/(4\sqrt{2}) = 0.146 \dots$ , where  $A_\nu$  can be computed as  $\int_{y=-\infty}^{\infty} \phi(\sqrt{2}y) [\Phi(y)]^{\nu-1} dy$ , or as  $g_\nu(1)$ . The function  $g_\nu(\alpha)$  is defined as  $g_1(\alpha) = 1/\sqrt{2}$ ,

$g_2(\alpha) = 1/2 \sqrt{2} \, dg_\nu(\alpha)/d\alpha = [(\nu - 1)(\nu - 2)/2\pi] \{ \alpha / [(2 + \alpha^2)(1 + \alpha^2)]^{1/2} \} g_{\nu-2} [\alpha / (1 + \alpha^2)^{1/2}]$ ,  $\nu \geq 3$ , with  $g_\nu(0) = 2^{-(\nu-1/2)}$ . Furthermore, as  $\nu$  approaches infinity,  $A_\nu \sim (2/\nu^2) \sqrt{\pi \log \nu}$ , so that  $G_n$  approaches  $\pi \log 2 = 2.18 = 3.4$  db. With criterion  $p_W$ , there is an asymptotic power gain for  $n \geq 3$  as the signal-to-noise ratio approaches zero. This asymptotic power gain approaches 3.4 db from below as  $n$  approaches infinity.

This result can be quickly adapted to the criterion  $p_B$ , using the remark from Ref. 2, Sec. V, that in an orthogonal code, given that an error is made, all error patterns have the same probability. Thus the following theorem holds.

*Theorem 3:* With  $p_B$  as the criterion, there is a power loss, rather than a gain, using orthogonal codes with correlation detection as the signal-to-noise ratio approaches zero. This loss is asymptotic in  $n$  to  $(\pi \log 2) n^2 / 2^{2n}$ , or in db, asymptotic to a loss of  $(2n \log_{10} 2)$  db, an arbitrarily large loss, proportional to  $n$ .

*Proof:* Since every error pattern is equally likely, given that an error has been made, one finds, as in Ref. 2, that

$$p_B = \left( \frac{2^n - 1}{2^n - 1} \right) p_{WD}. \quad (44)$$

Substituting for  $p_{WD}$  in Eq. (38) yields

$$p_B(s) = \frac{1}{2} - \left( \frac{n}{2\pi} \right)^{1/2} \frac{1}{2^n - 1} A_{2^n - 1} s + O(s^2). \quad (45)$$

With no coding, on the other hand,

$$p_B(s) = \frac{1}{2} - \frac{s}{\sqrt{2\pi}} + O(s^2). \quad (46)$$

The asymptotic power gain is thus

$$\left[ \frac{n}{2^{2(n-1)}} \right] A_{\nu}^2 \frac{2}{2^n - 1}. \quad (47)$$

This ratio is always less than 1, since  $A_{\nu}$  is always less than 1. The remainder of the theorem is completed using Eq. (37).

*Remark:* Correlation detection is, as before, not the best way to decode to minimize  $p_B$ . The same remarks apply here as in the bit-by-bit detection case: a given information symbol is called 0 if 0 is the most likely symbol in that position, given the received waveform.

In summary, if the criterion is  $p_B$ , coding at low signal-to-noise ratios actually causes a loss of power; if  $p_W$  is the criterion, bit-by-bit codes barely help, but orthogonal codes give a gain of 3.4 db at low signal-to-noise ratios.

## REFERENCES

1. Hackett, C. M., Jr., "Word Error Rate for Group Codes Detected by Correlation and Other Means," *IEEE Transactions on Information Theory*, Vol. IT-9, 1963, pp. 24-33.
2. Viterbi, A. J., *On Coded Phase-Coherent Communications*, Technical Report No. 32-25, Jet Propulsion Laboratory, Pasadena, California, 1960.
3. Peterson, W. W., *Error-Correcting Codes*, John Wiley and Sons, Inc., New York, 1961.
4. Feller, William, *An Introduction to Probability Theory and its Applications*, 2nd ed., John Wiley and Sons, Inc., New York, 1957.